

● ● ● 測定の目的

- 追跡のための情報
 - 対象者、友人、親類などの連絡先
- ランダム化の調整因子
- 研究の一般化のための情報
- 群間の比較可能性のための情報

13

● ● ● 測定の目的

- 追加解析
 - サブグループ解析、交絡を調整した解析の説明変数
 - 治療開始前のデータを用いることが大事
 - 結果因子も測定しておく

14

● ● ● 目次

- 対象者の選択
- ベースラインデータの測定
- ランダム化
- 介入の実施

15

● ● ● ランダム化

- ランダム化、ランダム割り付け
- Randomized、Random allocation、Random assignment
- ここでは「ランダム化」に統一

16

● ● ● ランダム化とは

- 予見不可能に群を決定すること
 - Selection biasを防ぐ
- 群間の比較可能性を確率的に保証する
 - 未知の因子についても保証する
 - 偶然偏ることもある(Accidental bias)
- 第三者が割り付けるのが理想
- ランダム化の遵守
 - 封筒法

17

● ● ● ランダム化の前に

- ベースラインデータの測定
- 選択・除外基準を満たすことの確認
- インフォームドコンセントの確認

18

● 小規模研究におけるランダム化

- コイントス
- 乱数の利用
 - 乱数表やコンピュータでの乱数発生
 - 1, 9, 3, 6, 4, ...
 - 偶数なら治療群、奇数ならコントロール群
 - 確率 p の二項乱数(ある確率 p で1/0が出る)
- 予見不可能性の保証が大切
 - × ABABAB...

19

● 大規模研究におけるランダム化

- 第三者機関で割付
 - 登録センター、事務局
 - 被検者がエントリーするたびにFAX、メールなどで連絡
 - センターで割付、結果を返信
- 研究者の意図が入らないことが大事

20

● ランダム化の方法

- 割付比
- ランダム化法
 - 通常のランダム化(Fixed Randomization)
 - 動的割付(Adaptive Randomization)

21

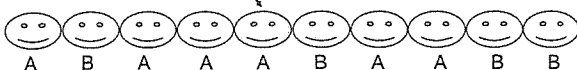
● 割付比

- 均等割付
 - 実薬群: プラセボ群=1:1
- 不均等割付
 - 各群の人数比をR:1で割り付ける
 - 割付比2:1で検出力がやや上昇、感度は減少(Peto 1978)
 - 複数の介入群と比較されるコントロール群を多くする
 - Uncertainty, Equipoiseの原則に反する
 - ICが難しくなる
- 均等割付が望ましい

22

● シンプルランダム化

- ランダムにAとBを発生させる
 - A・B・A・A・A・B・A・A・B・B
 - Aが多い(6:4)
 - 全員がAの属性がある



23

● 偏りを防ぐ方法

- Fixed Randomization
 - ブロックランダム化
 - 層別ランダム化 (≠クラスターランダム化)
- 動的割付(Adaptive Randomization)
 - 最小化法
 - つぼモデル
 - その他

24

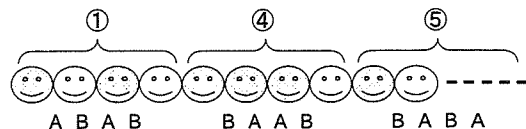
● ● ● ブロックランダム化

- A,Bそのものではなく「割付ブロック」を用いる
- 治療群をブロック内で均等に配列
- 例:ブロックサイズ=4
 - ① ABAB
 - ② ABBA
 - ③ AABB
 - ④ BAAB
 - ⑤ BABA
 - ⑥ BBAA

25

● ● ● ブロックランダム化

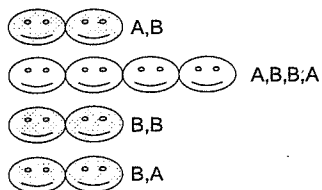
- 4人に①～⑥を割り付ける
- 4人おきに各群の人数が均等になる
- エントリー時期による対象者の質の変動にも対応



26

● ● ● 層別ランダム化

- 均等にしたい因子で層を作る
- 年齢グループで層を作ってランダム化



27

● ● ● 一般的に

- 「層別ブロックランダム化」が多い
- 層別因子
 - 変動の少ない、重要なものを2~3個
- ブロックサイズ
 - 治療群の数の倍数
 - 4~6が多い
 - サイズは混ぜることもある
 - プロトコルに書かない

28

● ● ● クラスターランダム化

- 複数人の集団をランダム化の単位とする
 - 例:施設ごとにランダム化
 - A病院(患者数 ≥ 1)に実薬
 - B病院にプラセボ...
- 用いる場面
 - 人数が少ない施設がたくさんあるとき
 - 介入が複雑なとき
- 効率は落ちる
 - より多くのサンプルサイズが必要

29

● ● ● 動的割付

- 割り付け結果を見ながら次の人を割り付ける
- サンプルサイズに対して均等にしたい因子が多い場合に有用
- 層別割付より若干検出力が高い
- 最小化法
- つぼモデル

30

● ● ● 最小化法(Minimization)

- それまでの割付結果に応じて割付確率を変える

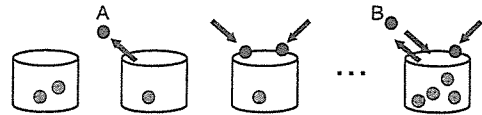
		A治療	B治療
性別	男性	20	19
	女性	10	11
重症度	I、II	19	20
	III、IV	11	10

- 次の患者: 男性、III
 - A: $21 + 12 = 33$
 - B: $20 + 11 = 31$
- \Rightarrow 偏りが事前に設定した閾値を越えたら Bに入る確率を高くする ($P=2/3$ など)

31

● ● ● つぼモデル(Urn model)

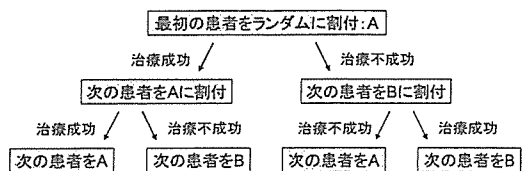
- 2種類の玉を復元抽出
 - 赤がでたらAを割り付け、青を足す
 - 青が出たらBを割り付け、赤を足す



32

● ● ● その他の動的割付

- 前の対象者の治療結果に基づくランダム化
Response adaptive randomization/Play the winner
- その時点でよりよいと思われる治療に割り付ける



33

● ● ● その他の動的割付

- Equipoiseの仮定から見れば妥当
- バイアスを招く可能性が高い
- 使える状況に限られる
 - 治療結果の出現が早いとき
 - 重要な結果因子が1個に決まっているとき
 - エントリー時期に関わらず患者集団の質が変わらないと仮定できるとき
- 他の方法より検出力は損なわれる
- 一般的ではない

34

● ● ● 目次

- 対象者の選択
- ベースラインデータの測定
- ランダム化
- 介入の実施

35

● ● ● 介入の実施

- Blinding (Masking、盲検化)
- 介入の選択
- コントロールの選択

36

● ● ● 盲検化

- 試験を実際に行う人が割付結果を知らないこと
- 比較治療同士が見分けがつかないことが前提
- 研究過程でのバイアス混入を防ぐ
 - 割付の遵守におけるバイアス
 - 重症な患者に新治療を割り付ける
 - プラセボ群の被験者が他の薬に頼る
 - 結果の評価におけるバイアス
 - よくなった気がする
- ランダム化は研究開始時点でのバイアスを防ぐ手順

37

● ● ● 盲検化のレベル

- Un-blinding(非盲検)
- Single Blinding(単盲検)
 - 被験者のみが知らない
 - 併用治療や評価においてバイアスがりうる
- Double Blinding(二重盲検)
 - 臨床医も被験者も知らない
- Triple Blinding(三重盲検)
 - データモニタリング委員会もどっちが実薬か知らない
 - 安全性に問題がある

38

● ● ● 盲検化のコスト

- 盲検化に必要なもの
 - プラセボの製造
 - コード番号の貼付
 - 記録の保持
 - コード番号のみでの配送、配役
 - 薬剤部との連携
 - 緊急時のためのKey

39

● ● ● 盲検化の維持

- 検査結果を誰が管理するか
 - 治療者、被験者に検査結果を知らせない
 - ダミーデータを用いる
 - 基準値を超えたときはそれだけを伝える
- 被験者に対する推量テスト
 - バイアス混入の可能性について議論する

40

● ● ● 盲検化できないとき

- わかってしまう薬もある
 - におい、色、味、注射跡、痛み...
- PROBE試験
 - Prospective Randomized Open Blinded-Endpoint
 - 結果の評価だけは盲検的に行う
- できるだけ「ハード」な結果指標が望ましい

41

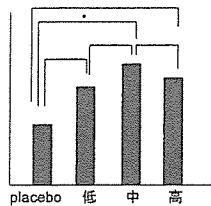
● ● ● 介入の選択

- 効果と安全性のバランスのとれた用法用量
- 研究対象となる疾患によって異なる
 - 重篤な疾患、致死率の高い疾患の治療
 - 治療効果優先
 - 最大許容量(highest tolerable dose)
 - 重篤度の低い疾患の治療、予防的治療
 - 安全性の確保を重視
 - 最低効果量(lowest effective dose)

42

効果と安全性のバランス

- 用量が定まらないときは複数群(低・中・高用量)で行うこともある
- ただし検定の多重性が生じる
 - 同じ仮説を何度も検定をすると本当は差のない仮説のどこかに偶然有意差がでてしまう確率(α エラー)が大きくなる
 - 有意水準を変える
 - $P < 0.05/6$



43

用量設定のための試験

- 有意差にこだわらず広範囲の用量や用法を試す
- FDAでいう第II相試験
 - 真のエンドポイントでは危険、時間がかかりすぎる
 - サロゲートエンドポイントを用いる
 - 真のエンドポイントとの関連を確認
 - 複数の指標を調べる
 - 例: TC, LDL, HDL, 血圧, BMI...
- 安全性を調べる

44

治験では

	目的	デザイン
第I相	安全性の検討 薬物動態の探索	健常人対象 並行ランダム化 プラセボを用いた単盲検 10~40例
第II相	有効性の確認 用法用量の決定 安全性の検討	患者対象 並行ランダム化 プラセボを用いた二重盲検 100~500人
第III相	有効性の検証 安全性の検討	並行ランダム化 二重盲検 プラセボ対照優越性 実薬対照優越性/非劣性

第I相試験って...



46

多剤併用する場合

- 効果の原因が特定できない
 - 「エストロゲン+プロゲステロン」で効果があった
 - どちらの薬剤の効果?
 - 単剤では効かない?
 - 単剤で行われた先行研究が必要
 - 要因実験 (Factorial design)
- 介入はシンプルのほうがよい
 - 実施可能性、コスト、解釈の点で

47

コントロールの選択

- 理想のコントロール
 - 「治療を受けた人」が、治療を受けずにそれ以外の点で全く同じ状態にいる場合
- 現実的に、理想のコントロール
 - 治療群と背景が似ている集団
 - ランダム化に基づく
 - 治療を受けていない盲検下の集団
 - 実薬とそっくりなプラセボを服用している
 - 盲検によってバイアス(治療を受けていないという自覚)が除かれる

48

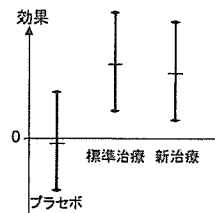
● ● ● プラセボが使えない場合

- 倫理的に不可能な場合
- 共介入(co-intervention)がある場合
- 治療効果の過小評価になりがち
 - 情報を取っておき、副次的に調整した解析を行う
- すでに標準治療がある場合
 - 試験デザインを変える

49

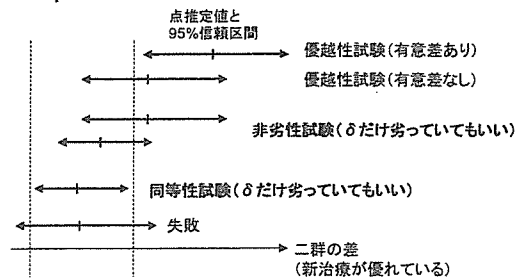
● ● ● 標準治療がある場合

- プラセボ治療と比較するのは非倫理的
- 治療効果は標準治療と同じくらい
- 他の利点がある
 - 副作用が少ない
 - 服用回数が少ない
 - 安価
- 治療効果で優越しなくてよい
- 非劣性試験、同等性試験



50

● ● ● 優越性、同等性、非劣性



大橋CE12回より 51

● ● ● まとめ

- 対象者の選択
- ベースライン測定
- ランダム化
- 盲検化
- 効果と安全性のバランス
- コントロールの選択

52

● ● ● 参考文献

1. Hulley SB, Cummings SR, Browner WS, et al. 木原雅子ら訳. 医学的研究のデザイン 第2版. MEDSI. 2004.
2. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials - 3rd Chap5, 6. New York. Springer. 1996.
3. Piantadosi S. Clinical Trials 2nd. Chap13. New Jersey. Wiley. 2005.
4. Green S, Benedetti J, Crowley J. 福田治彦ら訳. 米国SWOGに学ぶ 癌臨床試験の実践. 医学書院. 2004.
5. 大橋靖雄, 荒川森弘. 臨床試験の進め方. 南江堂. 2006.
6. 大橋靖雄. Critical Eyes on Research Methodology -がん研究のための方法論第12回
7. 佐藤俊哉. ヘルスサイエンスのための統計科学連載2「治療のランダム割付と治療効果の検定」<http://www.kbs.med.kyoto-u.ac.jp/ayumi1.pdf>

53

● ● ● 今後の予定

- 臨床試験II(治療効果とITT)
- 生存時間解析
- サンプルサイズ設計
- QOL
- 中間解析とIDMC
- 第I相、II相
- メタアナリシス
- ベイズ流解析
- 代替エンドポイント

54

医学的検査に関する研究をデザインする

- 因果関係よりも、診断検査が臨床的に有用かどうかの評価が主。
- 有意差検定よりも、感度sensitivity、特異度specificity、信頼区間confidence intervalなどの記述統計が用いられる。
- 有用性の評価には、再現性reproducibility(定度(精度?)precision)、真度accuracy(正確性)、実施可能性feasibility、臨床判断clinical decision、アウトカムに及ぼす効果などを検討する必要がある。

医学的検査における一般的事項

- 患者のスペクトルと検査結果
対象者は、実地臨床で遭遇するような患者や非患者を反映するようなスペクトルをもつものでなければならない。患者が重症者に偏ったり、コントロール群が特別健康な人に偏ったりしない。
- 盲検化の重要性
判定者に、検査以外の臨床情報を与えない。
例) 虫垂炎の超音波診断者に、患者の病歴や身体条件を知らせない。

測定の再現性に関する研究

- 観察者内変動 intraobserver variability
(例) 骨年齢のTW2法による測定
同じレントゲンを、間隔をあけて読影する
A: 10.3歳、10.7歳、10.5歳、10.1歳、10.6歳
mean±SD 10.44±0.24歳
変動係数 Coefficient of variation
標準偏差÷平均×100(%)=2.3%

測定の再現性に関する研究

- 観察者間変動 interobserver variability
(例) 骨年齢のTW2法による測定
同じレントゲンを、数名で読影する
A: 10.3歳、B: 10.7歳、C: 11.5歳、D: 9.6歳、E: 10.6歳
mean±SD 10.54±0.69歳
変動係数 Coefficient of variation
標準偏差÷平均×100(%)=6.5%

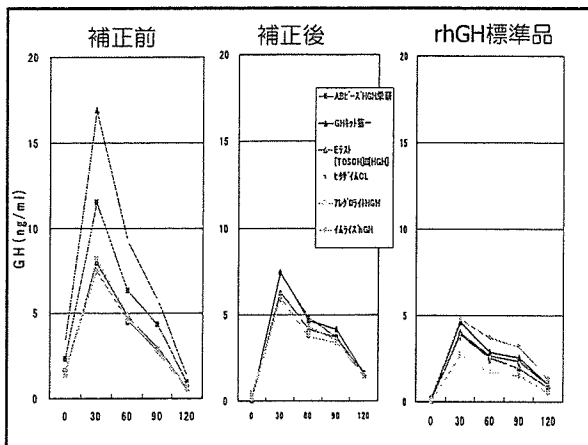
測定の再現性に関する研究

- 測定内変動 intra-assay variability
同じ検体を、1つ測定法(キット)で、何回も測定する。
(例) 成長ホルモンの測定
1つの検体をAキットで測定したとき。
10.2ng/ml, 10.3ng/ml, 10.2ng/ml
10.23±0.06ng/ml CV 0.5%
- しかし、同じキットでも別の施設で測定したときは、施設間のinter-assay variabilityを見ることになる。

測定の再現性に関する研究

- 測定間変動 inter-assay variability
同じ検体を、異なる測定法(キット)で、測定する。
(例) 成長ホルモンの測定
2002年成長科学協会 GH関連因子検討委員会

	栄研	第一	東ソー	日立化成	ヤマトロン	協和	平均	SD	CV
0分	0.12	0.15	0.0	0.02	0.31	0.1	0.12	0.11	92.8
15分	0.58	12.84	7.2	3.14	4.50	6.6	7.32	3.49	47.8
30分	13.25	18.80	10.9	6.89	5.85	9.6	10.82	4.63	42.3
60分	11.05	13.66	8.9	5.97	5.40	6.5	8.63	3.25	38.8
60分	6.27	9.91	6.2	4.02	3.15	4.2	5.47	2.14	39.2
120分	2.25	3.57	2.2	0.91	0.70	1.8	1.91	1.05	54.9



測定の再現性に関する研究

・ 日本アイトープ協会 全国コントロールサーベイ

キット名	試料A-1				試料A-2			試料A-3		
	N	Mean	± SD	CV	Mean	± SD	CV	Mean	± SD	CV
DPG-ENRGA	16	29.15	± 4.13	10.5	165.61	± 11.56	7	282.2	± 28.40	10.2
912AACS-E2-421-99AAS (GLIA)	10	84.85	± 7.85	9.2	177.28	± 9.71	5.5	244.5	± 15.33	6.2
912AACS-E2-411 (GLIA)	8	85.89	± 6.21	7.1	162.46	± 7.86	4.8	274.3	± 12.55	4.6
E2-3177 (PA)	7	47.2	± 6.97	14.7	212.26	± 28.68	13.5	811.2	± 117.75	14.5
S2A-2AEE (GLIA)	6	104.11	± 9.87	9.5	287.43	± 16.64	5.8	597.2	± 28.04	4.7
4mE24747 (GLIA)	3	107.01	± 11.2	10.5	242.10	± 21.65	9.0	472.2	± 128.82	27.0

測定の再現性に関する研究

- ・ 観察者変動の原因となる検査ステップをどこまで明らかにするかは、研究の目的による。
(例) 骨年齢の読影の差
どの骨の評価が異なるか
- ・ 再現性はprecision(測定結果が安定である割合)に関する概念であり、真度accuracyを保証しているわけではない。

測定の真度(正確性)に関する研究

- ・ ゴールドスタンダード(または定義によるゴールドスタンダード)が必要で、研究する検査が、どれだけゴールドスタンダードに一致するかを検証する。
- ・ ケースコントロール研究、横断研究、コホート研究など

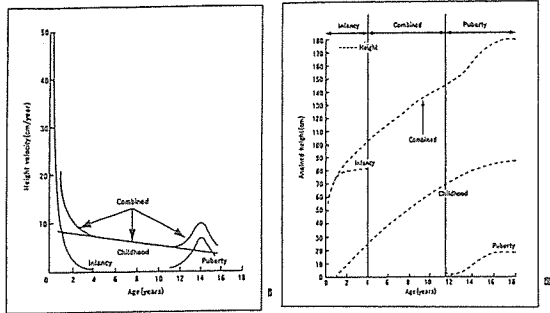
測定の真度(正確性)に関する研究

- ・ 感度sensitivity: 検査結果が陽性であった人の中で、疾患を実際に有していた人の割合。
- ・ 特異度specificity: 検査結果が陰性であった人の中で、疾患を有していなかった人の割合。

測定の真度(正確性)に関する研究

- (例) 3歳から6歳までの身長SDSの低下の、成長ホルモン分泌不全性低身長症(GHD)の診断に対する有用性。Retrospectiveな予後判定研究。
背景: GHDは、5~6歳に診断して早期治療するのが、短期的治療効果(早期の身長SDSの改善)、長期的治療効果(成人身長)がよいと言われている。GHD診断のためのGH分泌刺激試験を行う低身長児を選択する検査として、3歳から6歳までの身長SDSの低下を評価する。
- ・ 多くの低身長児は、3歳頃までに低身長になるが、GHD児はそれ以降に身長SDSの低下が著明になる。

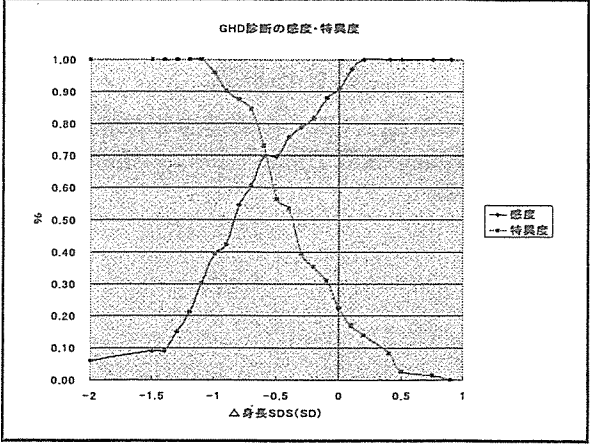
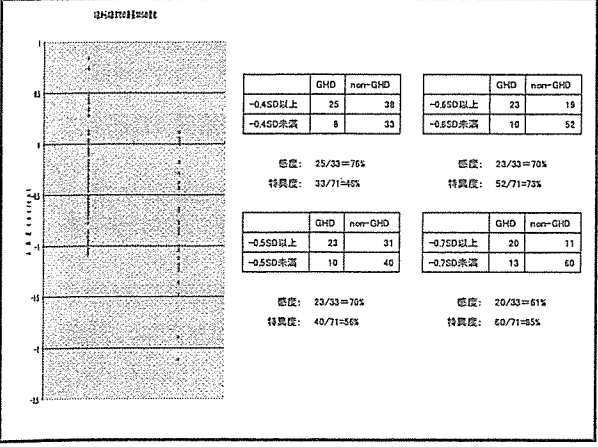
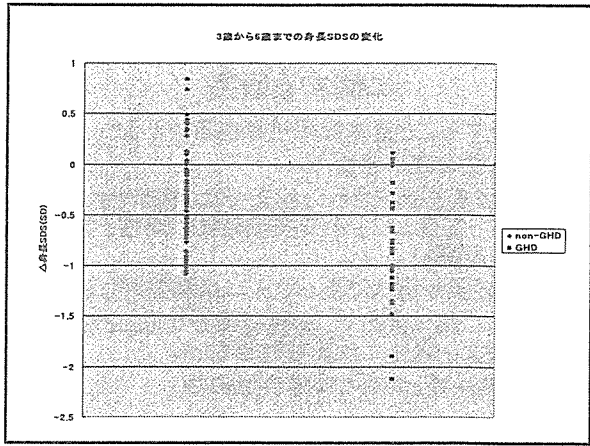
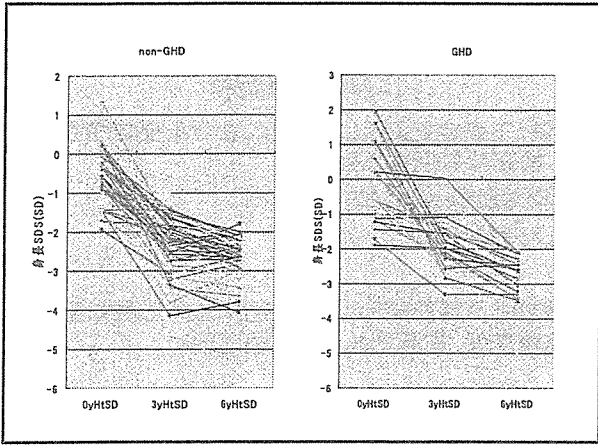
ICPモデル

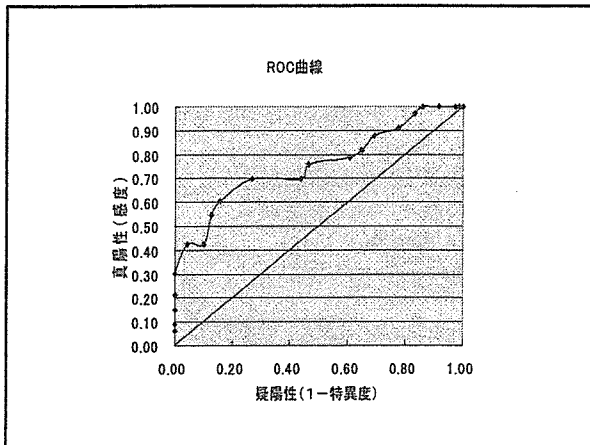


測定の真度(正確性)に関する研究

- 対象: 国立病院を受診した低身長児136名。
Including criteria
6歳時に身長SD: $-2SD$ 以下
出生時と3歳時の身長測定有り。
- Exclusion criteria
出生時身長SD: $-2SD$ 以下

104名が最終的な対象で、そのうち33名が2つ以上のGH分泌刺激試験でGHDと診断されている。
.....ゴールドスタンダード





尤度比 (likelihood ratio)

△身長SDS	GHD	(%)	non-GHD	(%)	尤度比
<-1.0	13	41.9	3	4.2	9.9
-1.0~-0.6	10	32.3	16	22.5	1.4
-0.6~-0.2	4	12.9	27	38.0	0.3
-0.2~0.2	6	19.4	15	21.1	0.9
>=0.2	0	0.0	10	14.1	0.0

臨床判断における検査結果の有用性に関する研究

- 検査すること自体有用性があるかどうか
- 実益性の高い検査
 - (1) 検査を指示された検査では、指示されなかった患者よりもその検査結果が陽性となる確率が高い
 - (2) 結果が陰性であった人には利益がない
- 非常に稀な疾患で、陽性結果がでることが滅多にない場合には実益性がない。
(最終診断と結びつく場合は必要)

臨床判断における検査結果の有用性に関する研究

(例) 下痢症で入院した患者における便の培養検査の有用性 (Siegel)

陽性結果: 40/1964例 (2%)
入院後3日以上: 0/997例

(結論) 入院後3日以上 of 患者における便培養検査は、有用である可能性は低い。

臨床判断における検査結果の有用性に関する研究

- 陽性になることが多いが、病歴や身体検査、またその他の検査で得られる以上の情報が得られない。

診断検査の実施可能性、コスト、リスクに関する研究
検査実施がアウトカムに及ぼす影響に関する研究

- 偽陽性結果の医学的・心理的影響の研究など
- (例) クレチン症スクリーニングで偽陽性になった新生児の親に、子供の健康に関する不安が半年続く
- 検査が正しい場合でもQOLの低下をきたす
- (例) 骨密度が異常といわれた女性では、運動を制限する傾向がある。

スクリーニング検査

- 偽陽性がすくない
- 陽性になった人に対する、治療法が確立されていて、確実に利益が増える
- コスト・ベネフィット

- 新生児マススクリーニング
- 低身長は、スクリーニングの対象にならない

スクリーニング検査

- S字結腸内視鏡検査が大腸ガンの死亡率の減少につながるか
- コホート内ケースコントロールスタディー
- 大腸ガンで死亡した患者とそうでない患者の内視鏡検査受診頻度 オッズ比0.41(95%信頼区間=0.25~0.69)60%の死亡率の減少
- 受けた人がガンになりにくい可能性
- 近位結腸ガン(内視鏡ではわからない)の死亡率 オッズ比0.96(95%信頼区間=0.61~1.50)

スクリーニング検査

- 便の潜血検査によるスクリーニング
- 実際にそれによって発見された大腸ガン患者の生存率が、スクリーニングされたなかった人より高くなければ意味がない。
- 胸部XP
- 尿中VMA: 自然に治癒するような症例までもみつけて、治療していた。

診断検査研究のデザインや分析における留意点

- サンプルサイズの不足
陽性率が低い場合(1%以下)に、検査が高価でなければ、サンプルサイズを増やしてのその実施を正当化できる。
また、治療可能な重篤疾患の場合も、十分正当化される。

診断検査研究のデザインや分析における留意点

- 除外基準の不適切な適用
- 効果の施設固有性
- 境界例や解釈不能例の除外
境界領域の結果を1つのカテゴリーとして分類する必要がある。
検体の紛失など、患者自身によらない場合は除外できる。

実験的研究をデザインする 臨床試験Ⅱ

Designing Clinical Research 2nd Chap11
Mar2007臨床研究センター勉強会
国立成育医療センター臨床研究フェロー
斉藤真梨

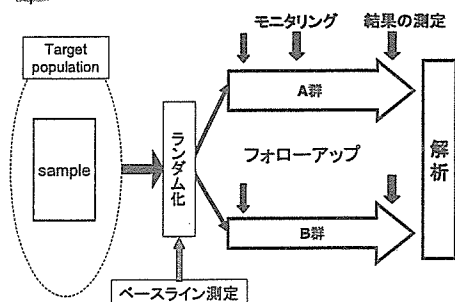
1

前回の内容

- 対象者の選択
 - Entry criteriaとExclusion criteria
- ベースラインデータの測定
 - 基本的属性、結果因子
 - 目的:ランダム化因子、追跡情報、背景の記述
- ランダム化
 - Selection biasを防ぐ
- 介入の実施
 - ブラインドと適切な用量設定

2

臨床試験のながれ



3

目次

- アドヒアランスとフォローアップ
- エンドポイント
- 結果の解析
- 脱落とその取り扱い

4

アドヒアランスとフォローアップ

- アドヒアランスの問題
 - 割り付けられた治療と違う治療を受けた
 - 途中でやめた
 - 併用禁止薬を用いた
- フォローアップの問題
 - 追跡中に受診しなくなった
 - 治療が続けられなくなった
- 悪影響
 - 余分にサンプルサイズが必要
 - 効果の推定にバイアスが入る
 - 評価できる効果が小さくなる

5

アドヒアランス

- 遵守度のこと
 - どれだけ試験プロトコルが守られているか
 - 医師、対象者にかかわらず
- 研究の質にかかわる
 - アドヒアランスが低い研究は結果の信頼性も低いと思われる

6

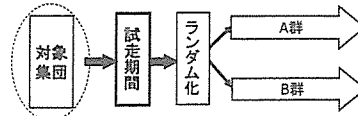
アドヒアランスの工夫

- アドヒアランスを高める
 - 受診日を対象者にあわせて調整する
 - 決まった時間に内服してもらう
 - 薬を小分けにする
- アドヒアランスを確認する
 - 日記などにより自己申告してもらう
 - 残った薬の数を数える
 - 尿中や血中の濃度を測定する

7

事前に確認

- ランダム化前に対象者を選ぶ
 - ためし受診(screening visit)
 - 試走期間(run-in-period)
 - 結果の一般化可能性は損なわれる
 - × ランダム化後に行うと比較可能性が崩れる
 - × 副作用の過小評価、漏れ



8

フォローアップの工夫

- 教科書158頁の表11-1
- アドヒアランスが高そうな患者を選ぶ
- 対象者が受け入れやすいように介入を工夫する
- 測定を痛みがないようにする
- 参加者に重要性を理解してもらう
- 試験を続けられるよう支援する
 - 連絡をまめに取る
- 脱落した人を探す
- 7章のコホート研究と同じ

9

フォローアップの工夫

- できるだけフォローアップを続ける
 - 「治療中止」
 - プロトコル違反があっても試験治療を中断しても可能な限り評価を続ける
 - 「試験中止」
 - 治療も追跡もやめてしまう
- 人を対象とした実験
 - 事前の計画を重視する

10

前回のHERS研究では

- 心血管疾患の既往のある女性2763人を追跡
 - エストロゲンとプロゲステロンの併用補充療法
 - 新規心血管疾患の発症
- ほぼ全員の結果を確認
 - 89%が最終受診
 - 8%は電話で
- 生存死亡情報は全員確認
 - 電話・書留郵便・近親者との接触・調査会社への依頼

11

KLIS

- Kyushu Lipid Intervention Study
 - 日本初のプラバスタチンを用いた大規模臨床試験
 - TC, LDL-Cの低下による冠動脈イベントおよび脳梗塞の初発抑制効果が示された
- 対象:
 - 高コレステロール血症
 - 45~74歳、心筋梗塞、脳梗塞の既往のない男性
 - 空腹時血清TC \geq 220mg/dL, HDL-C $<$ 80mg/dL
 - 除外基準: 既往歴、および2次性高脂血症、悪性新生物のあるもの

12

KLISのデザイン

- エンドポイント
 - 冠動脈疾患死、非致死性心筋梗塞、冠動脈再建術施行、突然死
 - 脳梗塞の発症および総死亡
- 非無作為化比較試験
 - プラバスタチン群3016人、従来治療群2579人
 - 封筒法がうまくいかなかった

13

KLISの結果

- 5640例登録
- 最終解析対象者3853例
 - 同意撤回147例
 - 厚生省の臨床試験基準の変更に伴う除外616例
 - 除外規定違反97例
 - 投薬違反、登録違反、データ欠損などによる除外223例
 - 残り700人は不明

14

エンドポイントの特性

- 精度(precision)と正確性(accuracy)
- 連続変数と2値変数
- 結果変数の数
 - たくさんの項目を評価することで結論を強固にできる
 - 多重性の問題も生じる
- 通常は主要エンドポイントをひとつ決める
 - それを元に症例数も決定する

15

目次

- アドヒアランスとフォローアップ
- エンドポイント
 - エンドポイントの特性と種類
 - 真のエンドポイントと代替エンドポイント
- 結果の解析
- 脱落とその取り扱い

16

エンドポイントの種類

- 主要評価項目
 - 一番興味があり、検証したい仮説に関する項目
 - 有効性に関する項目
 - 客観的な指標や妥当性が確保されている尺度が望ましい
- 副次評価項目
 - 有効性に関する評価項目
 - 違う項目
 - 違う解析
 - 安全性に関する評価項目

17

安全性評価項目

- 試験治療のよさはプラスの効果がマイナスの効果より大きいこと
- 有害事象
 - 臨床試験中に起こった良くない出来事の全て
 - 感冒症状や皮疹から死亡にいたるまで
- 副作用
 - 有害事象のうち試験治療と因果関係のあるもの
- 試験治療に起こりうるものを列挙し、程度を聞く

18

何を評価するか

- 真のエンドポイント(true endpoint)
 - 一番興味のあるもの
 - 死亡、心筋梗塞、入院、生活の質(QOL)など
 - 評価するには時間、お金がかかりすぎるものもある
- 代替エンドポイント(surrogate maker)
 - 結果と生物学的に妥当な関係がある
 - 結果と疫学的な関連をもつ
 - 心血管疾患に対する血圧、コレステロール
 - 骨折に対する骨密度

19

代替エンドポイント

- 結果が真のエンドポイントと逆転することもある
 - 原因不明ことが多い
- CAST (Cardiac Arrhythmia Suppression Trial)
 - 抗不整脈薬(エンカイド or フレカイド) vs プラセボ
 - 心筋梗塞発症後の患者1455人
 - 心室性不整脈を有意に抑制
 - 心血管系死亡が有意に上昇(33/730 vs 9/725)
 - 総死亡も有意に上昇(56/730 vs 22/725)

20

代替エンドポイントのよさ

- 真のエンドポイントと相関があること
- 疾患や真のエンドポイントによっても異なる
- 統計的にもいくつかの基準が提案されている
 - Prentice基準
 - PTE (Proportion of Treatment Effect)

21

統計的には

- Prentice基準
 - うけた治療によって代替エンドポイントの分布が異なる
 - うけた治療によって真のエンドポイントの分布が異なる
 - 代替エンドポイントの値が同じならどっちの治療を受けても真のエンドポイントは同じ
- PTE
 - 代替エンドポイントを経由する効果と治療の効果の比が1に近い



22

目次

- アドヒアランスとフォローアップ
- エンドポイント
- 結果の解析
 - 検定と推定
 - サブグループ解析
- 脱落とその取り扱い

23

結果の解析

- 検定
 - 片方の群が有意に優れているかに興味がある
 - 二群の比較
- 推定
 - 介入の効果の大きさをしりたい
 - 交絡を調整したい
 - 回帰モデル
 - 層別解析

24

検定

- 連続変数の比較
 - t検定 (結果指標が正規分布すると仮定)
 - ノンパラメトリック検定 (分布形を仮定しない)
 - 符号検定
 - ウィルコクソン検定 (順位にもとづく検定)
 - 並べ替え検定
- 割合の比較
 - χ^2 検定
 - Fisherの直接法 (人数が少ないとき)
- 時間の比較
 - 生存時間解析 (ログランク検定、Kaplan-Meier曲線)

25

効果の推定

- 効果を定量化する
- 交絡の影響を調整する
- 回帰モデル
 - 結果変数 = $a \times \text{リスク因子} + b \times \text{リスク因子} + \dots$
 - 興味ある因子の係数 a を推定する
 - リスク因子が1増加することに結果が a 増加する
- 層別解析
 - 背景因子ごとに興味ある因子の効果を算出し併合する

26

サブグループ解析

- 対象を特定の部分集団に絞って解析すること
 - 試験治療のターゲットを同定できる
 - テーラーメイド医療への示唆
- 例) 骨粗鬆症性骨折予防薬の臨床試験
 - 実薬群で14%のリスク減少
 - 介入前の測定で骨密度の低かった集団では36%のリスク減少
- 介入前のデータによって集団を特定
 - 部分集団でも比較可能性は維持される

27

サブグループ解析の欠点

- 人数が減るので有意差がでにくくなる
 - 有意差がでるように症例数設計していない
 - P値ではなく対象者数、効果の大きさ(リスク比、リスク差など)を結果として報告する
- 多重性が生じる
 - 様々な集団を作って解析すると偶然効果がある結果が出てしまうこともある

28

目次

- アドヒアランスとフォローアップ
- エンドポイント
- 結果の解析
- 脱落とその取り扱い
 - ITT解析、PP解析

29

ランダムに割り付けても

- プロトコル通りに実施されずとは限らない
 - 治療内容
 - 試験薬 ⇄ 対照薬
 - 併用薬を飲んでた
 - 治療期間
 - 途中でやめた、再開した
 - 途中で違う治療に変えた
 - データが測定できなかった

30

違反、欠測にはわけがある

- ランダムな違反、欠測
 - 結果と独立に生じる
 - 引越した
 - 交通事故にあった
- ランダムでない違反、欠測
 - 結果と関連がある
 - 副作用で中断した
 - よくなったからやめた
 - 悪くなったから他の治療に変えた

31

欠測のあるデータ

- 除いて解析するとバイアスが入る
 - 治療効果の過大／過小評価
 - 安全性情報の不足
 - ランダム化による群間の比較可能性が崩れる

32

KLISの結果

- 相対リスク低下
 - 冠動脈イベント14%低下
 - 脳梗塞22%低下
 - 冠動脈イベントと脳梗塞19%低下
 - 「初発抑制効果が示唆された」
- その他
 - TC低下: プラバ vs コントロール=15% vs 8%
 - LDL-C低下: プラバ vs コントロール=20% vs 10%

33

コンプライアンスによって

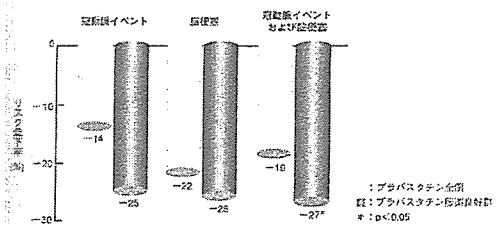


図 プラバスタチン治療による主要イベントのリスク低下率 (KLIS)

 本図はプラバスタチン治療により観察された結果を示す。
 観察指標: 全死, TC, LDL-C, HDL, 脳卒中, 冠動脈, 死亡率, 投与量, 投与量下位の効果, 効果, 効果
 The Pravastatin Intervention Study Group: J Atheroscler Thromb 7(2): 110-121, 1998 より引用

34

違反、欠測への対処法

- 割付どおりに扱う
- 違反した人は除外する
- 事前に想定すべきである
 - 予め除外基準をプロトコルに定めておく
 - その基準を満たさない場合のみ除外する

35

臨床試験では

- Intent-To-Treat (ITT)解析
 - 割付どおりに扱う
- Per Protocol (PP)解析
 - 評価可能例のみの解析
 - 違反した人は除外する
 - 事前に想定すべきである
- ITT、PPどちらでも治療効果があることを確認するのが理想

36

ITTの原則

- プロトコルの規定どおりに解析する
 - ランダム化による結果の妥当性が確保される
- 結果にどう影響するのか
 - 1) 治療効果がない場合
 - 2) 治療効果がある場合

37

架空の例

- 実薬群とプラセボ群でイベント発生割合を比較
- 実薬群とコントロール群の比較
- どちらの群もBが多いとコンプライアンスが悪い
- ITT解析
- PP解析

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

38

結果への影響

- 問題なのは治療効果の過大評価
 - 効果がないのにあると結論すること
 - 消費者危険
- ITTとPP、どちらがいいか
 - 治療効果がない場合
 - 治療効果がある場合

39

治療効果がない場合

- ランダムな脱落
 - 実薬でもプラセボでも発症する人は発症する
 - AでもBでも発症する人は発症する
- ITT、PPどちらもバイアスなく「効果なし」といえる

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

40

治療効果がない場合

- ランダムでない脱落①
 - 悪化した人が薬A→薬Bになった場合
 - 元気な人がプA→プBになった場合
- ITTでは「効果なし」といえる
- PPでは効果を過大評価

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

41

治療効果がない場合

- ランダムでない脱落②
 - 元気な人が薬A→薬Bになった場合
 - 悪化した人がプA→プBになった場合
- ITTでは「効果なし」といえる
- PPでは効果の過小評価

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

42

治療効果がある場合

- ランダムな脱落
- ITTでは効果の過小評価
 - 服用しなかった人はよくなる
- PPではバイアスなく評価できる

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

43

治療効果がある場合

- ランダムでない脱落①
 - 悪化した人が薬A→薬Bになった場合
 - 元気な人がプA→プBになった場合
- ITTでは効果の過小評価(実薬↓、プラセボ↑)
- PPでは効果の過大評価(実薬↑、プラセボ↓)

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

44

治療効果がある場合

- ランダムでない脱落②
 - 元気な人が薬A→薬Bになった場合
 - 悪化した人がプA→プBになった場合
- ITTでは効果の過小評価(実薬↓)
- PPでは効果の過小評価(実薬↓、プラセボ↑)

		実際の服用	
		守った	守らなかった
割付	実薬群	薬A	薬B
	プラセボ群	プA	プB

45

控えめな結果

- ITT解析
 - 「効果がない」とき効果がないといえる
 - 「効果がある」とき効果を過小評価する
- PP解析
 - 「効果がない」とき
 - ランダムな脱落なら効果がないといえる
 - ランダムでない脱落ならバイアスが入る
 - 「効果がある」とき
 - ランダムな脱落ならバイアスが入らない
 - ランダムでない脱落ならバイアスが入る

46

仮説検定の基本

- 検証試験では帰無仮説について議論する
- 「実薬には効果がない」
 - 効果がないとき帰無仮説を棄却しない
 - 効果があるとき棄却する
- どの脱落もランダムかどうかわからない
- ITTはバイアスの方向性が常に過小方向

47

まとめ

- アドヒアランスを高める工夫、高い人を選ぶ
- 治療をやめても追跡する
- 一番大事なエンドポイントを決める
- 結果の解析
- ITT解析の結果を優先する

48